

Sequence Alignment

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 22, 2004, 13:50:23 ; Search time 18 Seconds
(without alignments)
1423.251 Million cell updates/sec

Title: US-09-978-188A-7

Perfect score: 2527
Sequence: 1 MVKFPALTYWPLRFLVPL.....TMDPTEVTDIVEMRENE 492

Scoring table:
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	2527	100.0	492	1	ANKH HUMAN	Q9hcz1 homo sapien
2	2502	99.0	492	1	ANKH MOUSE	Q9hcz2 mus musculu
3	2489	98.5	492	1	ANKH RAT	P58366 rattus norv
4	2317	91.7	492	1	ANKH XENLA	P58367 xenopus lae
5	2158.5	85.4	501	1	ANKH BRARE	P58368 brachydanio
6	1528.5	60.5	355	1	ANKH TETNG	P58369 tetraodon n
7	114	4.5	416	1	LACY KLEOX	P18817 klebsiella
8	113.5	4.5	416	1	NUM FODAN	P15578 podospora a
9	111	4.4	622	1	COX1 BACSU	P24010 bacillus su
10	109.5	4.3	610	1	PT2A ARATH	P46031 arabidopsis
11	107	4.2	422	1	EXUT BACSU	O34456 bacillus su
12	107	4.2	533	1	LAT2 RAT	Q9wvr6 rattus norv
13	106	4.2	334	1	YB77 METJA	O58578 methanococc
14	105	4.2	401	1	LSG1 HAEIN	P71399 haemophilus
15	105	4.2	676	1	HPF1 METAC	O8tja9 methanosarc
16	103.5	4.1	398	1	BCR HAEIN	P45123 haemophilus
17	103	4.1	417	1	LACY ECOLI	P02920 escherichia
18	102.5	4.1	446	1	CITN SALTY	P31604 salmonella
19	102.5	4.1	531	1	LAT2 MOUSE	Q9qkw9 mus musculu
20	102.5	4.1	533	1	WVIN SHIR	O05467 rhizobium t
21	102.5	4.1	641	1	NUSM ALLMA	P50365 allomyces m
22	101.5	4.0	446	1	CITN SALDU	P31603 salmonella
23	100	4.0	582	1	YS96 CABEL	Q09965 caenorhabdi
24	99	3.9	385	1	Y421 METTH	O26521 methanobact
25	99	3.9	535	1	LAT2 HUMAN	Q9uh15 homo sapien
26	98.5	3.9	583	1	NUM NEUCR	Q35140 neurospora
27	98.5	3.9	946	1	YB7C YEAST	P38250 saccharomyc
28	98	3.9	480	1	YE44 YEAST	P37555 bacillus su
29	98	3.9	532	1	YAEW BACSU	P79235 pongo pygma
30	97.5	3.9	346	1	FMLR PONPY	Q818u7 drosophila
31	97.5	3.9	3803	1	TRAI DROME	P37482 bacillus su
32	97	3.8	402	1	YYCB BACSU	P79176 gorilla gor
33	96.5	3.8	346	1	FMLE GORGO	

ALIGNMENTS

RESULT 1	ANKH HUMAN	ANKH HUMAN	STANDARD;	PRT;	492 AA.
ID	ANKH HUMAN	ANKH HUMAN	STANDARD;	PRT;	492 AA.
AC	Q9HCJ1: Q9HCJ2;				
DT	28-FEB-2003 (Rel. 41, Created)				
DT	28-FEB-2003 (Rel. 41, Last sequence update)				
DT	10-OCT-2003 (Rel. 42, Last annotation update)				
DE	Progressive ankylosis protein homolog (ANK).				
GN	ANKH OR KIAA1581.				
OS	Homo sapiens (Human)				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
CL	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=20355194; PubMed=10894769;				
RA	Ho A.M., Johnson M.D., Kingsley D.M.;				
RT	"Role of the mouse ank gene in control of tissue calcification and arthritis";				
RL	Science 289:265-270(2000).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Brain;				
RX	MEDLINE=20450683; PubMed=10997877;				
RA	Nagase T., Kikuno R., Nakayama M., Hirose M., Ohara O.;				
RT	"Prediction of the coding sequences of unidentified human genes. XVIII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro.";				
RL	DNA Res. 7:273-281(2000).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Ovary;				
RX	MEDLINE=22388257; PubMed=12477932;				
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,				
RT	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,				
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,				
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,				
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,				
RA	Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,				
RA	Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,				
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mallaby S.J.,				
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,				
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,				
RA	Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,				
RA	Fahy J., Hulton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,				
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,				
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,				
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,				
RA	Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,				
RA	Schneerch A., Schein J.E., Jones S.J.M., Marra M.A.;				
RT	"Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences";				
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).				
RN	[4]				
RP	VARIANTS CMJG SER-375 DEL; PHE-376 DEL AND ALA-380 INS.				
RX	MEDLINE=21313103; PubMed=11326338;				

P47234 citrobacter
Q8d9n8 vibrio vuln
P58164 escherichia
P24975 balaeopter
P11982 asterina pe
P45272 haemophilus
P35865 corynebacte
Q9w3v8 drosophila
P50367 rhizopus st
Q91y73 arabidopsis
Q00758 bacillus su
P03660 bacterioph

RA Reichenberger E., Tiziani V., Watanabe S., Park L., Ueki Y.,
RA Santanna C., Baur S.T., Shiang R., Grange D.K., Beighon P.,
RA Gardner J., Hamersma H., Sellars S., Ramesar R., Lidral A.C.,
RA Sommer A., Raposo do Amaral C.M., Gorlin R.J., Mulliken J.B.,
RA Olsen B.R.,
RA "Autosomal dominant craniofacial dysplasia is caused by mutations
RT in the transmembrane protein ANK.",
RT Am. J. Hum. Genet. 68:1321-1326 (2001).
RN [5]
RN VARIANTS CMDJ ARG-292; ARG-331; SER-375 DEL; PHE-377 DEL; ALA-380 INS
RP AND ARG-389.
RX NUBERG P., Thiele H., Chandler D., Hoehne W., Cunningham M.L.,
RA Rittner H., Leschik G., Uhlmann K., Mischung C., Harrop K.,
RA Goldblatt J., Borochowitz Z.U., Kotzot D., Westermann F., Mundlos S.,
RA Braun H.-S., Laing N., Tinschert S.,
RA "Heterozygous mutations in ANKH, the human ortholog of the mouse
RT progressive ankylosis gene, result in craniofacial dysplasia.",
RL Nat. Genet. 28:37-41 (2001).
CC -1- FUNCTION: Regulates intra- and extracellular levels of inorganic
CC pyrophosphate (PPi), probably functioning as Ppi transporter.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -1- TISSUE SPECIFICITY: Found in osteoblasts from mandibular bone and
CC from iliac bone; not detected in osteoclastic cells.
CC -1- DISEASE: Defects in ANKH are the cause of craniofacial dysplasia
CC dysplasia Jackson type (CMDJ) [MIM:123000]. CMDJ is a rare
CC autosomal dominant skeletal disorder characterized by abnormal
CC bone formation and mineralization in membranous as well as
CC endochondral bones. Progressive thickening of the bones can cause
CC narrowing of cranial foramina and can lead to severe visual and
CC neurological impairment, such as facial palsy and deafness.
CC -1- SIMILARITY: BELONGS TO THE ANKH FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF274753; AAF88039.1; -
DR EMBL: AB046801; BAB13407.1; ALT_INIT.
DR EMBL: BC009835; AAH09835.1; -
DR EMBL: BC014526; AAH14526.1; -
DR EMBL: HGNC:15492; ANKH.
DR MIM: 605145; -
DR MIM: 123000; -
DR GO: GO:0016021; C: integral to membrane; IDA.
DR GO: GO:0019857; C: outer membrane; TAS.
DR GO: GO:0030504; P: inorganic phosphate transporter activity; IDA.
DR GO: GO:0005315; P: locomotory behavior; NAS.
DR GO: GO:0007626; P: regulation of bone mineralization; TAS.
DR GO: GO:0030500; P: regulation of bone development; NAS.
DR GO: GO:0001501; P: skeletal development; NAS.
KW Transport; Phosphate transport; Transmembrane; Disease mutation;
KW Deafness.
FT DOMAIN 1 85 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 86 106 POTENTIAL.
FT DOMAIN 107 131 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 132 152 POTENTIAL.
FT DOMAIN 153 158 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 159 179 POTENTIAL.
FT DOMAIN 180 189 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 190 210 POTENTIAL.
FT DOMAIN 211 326 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 327 347 POTENTIAL.
FT DOMAIN 348 350 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 351 371 POTENTIAL.
FT DOMAIN 372 403 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 404 426 POTENTIAL.
FT DOMAIN 427 429 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 430 452 POTENTIAL.

FT DOMAIN 453 492 CYTOPLASMIC (POTENTIAL).
FT VARIANT 292 292 W -> R (in CMDJ).
FT 331 331 /FTID=VAR 012192.
FT VARIANT 331 331 C -> R (in CMDJ).
FT 375 375 /FTID=VAR 012193.
FT VARIANT 375 375 Missing (in CMDJ).
FT 376 376 /FTID=VAR 012194.
FT VARIANT 376 376 Missing (in CMDJ).
FT 377 377 /FTID=VAR 012195.
FT VARIANT 377 377 Missing (in CMDJ).
FT 380 380 /FTID=VAR 012196.
FT VARIANT 380 380 P -> PA (in CMDJ).
FT 389 389 /FTID=VAR 012197.
FT VARIANT 389 389 G -> R (in CMDJ).
FT 78 78 /FTID=VAR 012198.
FT CONFLICT 78 78 N -> S (IN REF. 1).
SQ SEQUENCE 492 AA; 54240 MW; 44BEE9089BDEC6B CRC64;
Query Match 100.0%; Score 2527; DB 1; Length 492;
Best Local Similarity 100.0%; Pred. No. 7.8e-195; Indels 0; Gaps 0;
Matches 492; Conservative 0; Mismatches 0;
QY 1 MKFPPALTYWPLIRFLVPLIGITNIAIDFGEQALNEGIAAVKEDAVEMLASGLAYSLMK 60
DB 1 MKFPPALTYWPLIRFLVPLIGITNIAIDFGEQALNEGIAAVKEDAVEMLASGLAYSLMK 60
QY 61 FFTGMSDFKNGVLGVFVNSKRDRTKAVLCMVVAGATAAFTHTLIAYSDIGYIINKLHV 120
DB 61 FFTGMSDFKNGVLGVFVNSKRDRTKAVLCMVVAGATAAFTHTLIAYSDIGYIINKLHV 120
QY 121 DESVGSKTRAFLYLAAPFPMDAMANTHAGILLKHKYSFLVGCASISDVIAQVFFVAIL 180
DB 121 DESVGSKTRAFLYLAAPFPMDAMANTHAGILLKHKYSFLVGCASISDVIAQVFFVAIL 180
QY 181 HSHLECEPPLIPILSYMGALVRCCTTCLGYKNHDIIPDRSGPBGDARIMKLSF 240
DB 181 HSHLECEPPLIPILSYMGALVRCCTTCLGYKNHDIIPDRSGPBGDARIMKLSF 240
QY 241 WPLALILATORISRPVNLVFSRDLGSSAAEAVALTATVPVGHMPYGLTEIRAVY 300
DB 241 WPLALILATORISRPVNLVFSRDLGSSAAEAVALTATVPVGHMPYGLTEIRAVY 300
QY 301 PAFDKNPNKLVSTNTVTAHHKKTFTVCMALSLTLCFVCMALSLTLCFVCMALSLTLCFV 360
DB 301 PAFDKNPNKLVSTNTVTAHHKKTFTVCMALSLTLCFVCMALSLTLCFVCMALSLTLCFV 360
QY 361 PAFDELCTVPLRIEFPVPVTVTAHKTGMLTKTFLAPSVLRIRIIVLIASLVLPY 420
DB 361 PAFDELCTVPLRIEFPVPVTVTAHKTGMLTKTFLAPSVLRIRIIVLIASLVLPY 420
QY 421 LGVHGATLGVGSLLAGFVGESTMVAIAACVYRKQKKMENESATEGEDSAMTDPTEE 480
DB 421 LGVHGATLGVGSLLAGFVGESTMVAIAACVYRKQKKMENESATEGEDSAMTDPTEE 480
QY 481 VTDIVEMRENE 492
DB 481 VTDIVEMRENE 492
RESULT 2
ANKH_MOUSE
ID ANKH_MOUSE STANDARD; PRT; 492 AA.
AC Q9JH22; O35138; O35139;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Progressive ankylosis protein (Fn54 protein).
GN ANKH OR ANK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]

Db 2401 GGCTTGCTTTCCCTCGCCCTTCTCGAAGTGCATTAGAGCGAGTCAATGAGCATCC 2460

Qy 2461 TAACCTTGCACTTTAGCTTTTACAGTGAAGTCTTAAGTCTCATCGCATCTAA 2520

Db 2461 TAACCTTGCACTTTAGCTTTTACAGTGAAGTCTTAAGTCTCATCGCATCTAA 2520

Qy 2521 TGCCAGGTTGCTGTAGGTAACCTTTTGAAGTGAATATATACCTGCTTCTGCTATCTTA 2580

Db 2521 TGCCAGGTTGCTGTAGGTAACCTTTTGAAGTGAATATATACCTGCTTCTGCTATCTTA 2580

Qy 2581 GTCAATACTCTGGGTACAGGTAATTTGAGAAATGCTACTACGCTTCCCTCCACACCAT 2640

Db 2581 GTCAATACTCTGGGTACAGGTAATTTGAGAAATGCTACTACGCTTCCCTCCACACCAT 2640

Qy 2641 ACCATAAGCAAGACATTTTATAAGCATACACAGTCACTATGCTGCTCCCTGAAATA 2700

Db 2641 ACCATAAGCAAGACATTTTATAAGCATACACAGTCACTATGCTGCTCCCTGAAATA 2700

Qy 2701 ACCATTCGAAATCCATGAGTCAATATTTTCTAAGTATTTTGGAAAGCAGGTTTTT 2760

Db 2701 ACCATTCGAAATCCATGAGTCAATATTTTCTAAGTATTTTGGAAAGCAGGTTTTT 2760

Qy 2761 CCTTTAAAAAATTTATAGACAGGTTTCACTAAATTTAGTCAAGATTTCTAGACTGA 2820

Db 2761 CCTTTAAAAAATTTATAGACAGGTTTCACTAAATTTAGTCAAGATTTCTAGACTGA 2820

Qy 2821 AAGAACCTAAACAAAAAATTTTAAAGATATAAATATATATGCTGATATGTTATGTAAT 2880

Db 2821 AAGAACCTAAACAAAAAATTTTAAAGATATAAATATATATGCTGATATGTTATGTAAT 2880

Qy 2881 TTAATTTAGGCTATATACATTTCTTATTTTCCATTTTCCATTTTCAATAAATGCTCTATACA 2940

Db 2881 TTAATTTAGGCTATATACATTTCTTATTTTCCATTTTCCATTTTCAATAAATGCTCTATACA 2940

Qy 2941 AAAAA 2945

Db 2941 AAAAA 2945

RESULT 3

AB046801

LOCUS

DEFINITION Homo sapiens mRNA for KIAA1581 protein, partial cds.

ACCESSION AB046801

VERSION AB046801.1 GI:10047236

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE

1 (sites)

1. Nagase, T., Kikuno, R., Nakayama, M., Hirose, M., and Ohara, O. Prediction of the coding sequences of unidentified human genes. XVIII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro

JOURNAL DNA Res. 7 (4), 273-281 (2000)

MEDLINE 20450683

PUBMED 10997877

REFERENCE

2 (bases 1 to 3928)

1. Chao, O., Negase, T., and Kikuno, R. Direct Submission

AUTHORS

TITLE

JOURNAL

Submitted (03-AUG-2000) Osamu Ohara, Kazusa DNA Research Institute, Department of Human Gene Research, 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: cdna.ohara@kazusa.or.jp, URL: http://www.kazusa.or.jp/huge, Tel: 81-438-52-3913, Fax: 81-438-52-3914)

LOCATION/Qualifiers

1. .3928

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="fj05690"

/note="vector:pBluescriptII SK plus"

Gene 1. .3928

/gene="KIAA1581"

CDS <1..1640

/gene="KIAA1581"

/note="Start codon is not identified."

/codon_start=3

/product="KIAA1581 protein"

/protein_id="BABI3407.1"

/db_xref="GI:10047237"

/translation="APQPRGGAAGAAAPPPPAASVGRPARVPAPRLPAESPRGSRCLVQSPAGTAVKPEPLTHVPLIRFLVPLIGLITWIAIDRGEQALNRRGIAAVKEDAVEMLASGLATSLMKFTFGMSDFKNVGLVFNVSKRRTKAYLCMVVAGALAAVPHLLIAYSIDLYIINKLHVDESVGSTRRAFLYLAAPFMDAMANTHAGILKHKRIFLVLGGASISDYIAVFAVVAIILSHLECRPLIPILSLYMGALVCTTICLGYKNHIDIIIDRSYGLGDATIRKMLSPWPLALILATIRSVINLVSRDLGGSSAAEVAAILTA TYPGHMYPGLTEIRAVYPAFKNNPSNKLVSNTVTAHIIKKTFTVCMALSLTLC FVNPENVSKEIILDIIGVDPAPBLCVPLIRIFSPFPVTVRAHLTGWMLTKKT FVLPSSVLRIILVILASLVLPYLVGHGATLVGSLAGVPGVSTVVAIAACVYRKQKKMENSATGEGSDSMTDMPPEEVTDIVEMRENE"

ORIGIN

Query Match 99.7%; Score 2936.6; DB 9; Length 3928;

Best Local Similarity 99.9%; Pred. No. 0;

Mismatches 0; Mismatches 4; Indels 0; Gaps 0;

Matches 2939; Conservative 0;

Qy 1 CGCTGCCCCCTCGCCCTCGCCCTCGCCGAGAGTCCCTCGCGGAGAGAGTGTGTG 60

Db 80 CGCTGCCCCCTCGCCCTCGCCCTCGCCGAGAGTCCCTCGCGGAGAGAGTGTGTG 139

Qy 61 GGGTCAGCCACGCGGGGACTATGTTGAAATTCCTCGGCGCTCACGCACTACTGGCCCT 120

Db 140 GGGTCAGCCACGCGGGGACTATGTTGAAATTCCTCGGCGCTCACGCACTACTGGCCCT 199

Qy 121 GATCGGTTTGTGTGCCCCCTGGGCATCACCAATAGCCATCGACTTCGGGAGCGAGC 180

Db 200 GATCGGTTTGTGTGCCCCCTGGGCATCACCAATAGCCATCGACTTCGGGAGCGAGC 259

Qy 181 CTGGAACCGGGCATTTGCTGTCTCAAGGAGATGTCAGTTCAGATGCTGGCCACTACG 240

Db 280 CTGGAACCGGGCATTTGCTGTCTCAAGGAGATGTCAGTTCAGATGCTGGCCACTACG 319

Qy 241 GTCGGCTACTCCCTCATGAAGTTCTTTCAGGGTCCCATGAGTCACTCAAAAATGTGG 300

Db 320 GTCGGCTACTCCCTCATGAAGTTCTTTCAGGGTCCCATGAGTCACTCAAAAATGTGG 379

Qy 301 CTTGGTGTTCGTAACAG 360

Db 380 CTTGGTGTTCGTAACAG 439

Qy 361 AGGGGCCATCGCTCGCTCTTTCACACTGATAGCTTATAGTGAATTTAGGATCTACAT 420

Db 440 AGGGGCCATCGCTCGCTCTTTCACACTGATAGCTTATAGTGAATTTAGGATCTACAT 499

Qy 421 TATCAATAAATCACTGACCATGTCAGTTCGGTGGGAGAGAGAGAGAGAGAGAGAGAG 480

Db 500 TATCAATAAATCACTGACCATGTCAGTTCGGTGGGAGAGAGAGAGAGAGAGAGAGAG 559

Qy 481 GTACCTCGCGCTTTTCTCTTTCATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540

Db 560 GTACCTCGCGCTTTTCTCTTTCATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 619

Qy 541 AAAAAACAAAATACAGTTTCTGTTGGGAGTGTCTCAATCTCAGATGTCATGCTCAGGT 600

Db 620 AAAAAACAAAATACAGTTTCTGTTGGGAGTGTCTCAATCTCAGATGTCATGCTCAGGT 679

Qy 601 TGTCTTTTGTAGCCATTTTGTCTTTCAGTCACTGATGAGAGAGAGAGAGAGAGAGAG 660

Db 680 TGTCTTTTGTAGCCATTTTGTCTTTCAGTCACTGATGAGAGAGAGAGAGAGAGAGAG 739

Qy 661 GATCCTCTCTTTGTACATGAGGCGCATTTGTGCTGTGACCAACCTGTGCTGGGCTACTA 720

Db 740 GATCCTCTCTTTGTACATGAGGCGCATTTGTGCTGTGACCAACCTGTGCTGGGCTACTA 799

QY	721	CAAGAA	CATT	CA	CGAC	ATCAT	CCCT	GC	ACAG	AGTGC	CCCG	AGCT	GGGG	GAGAT	GC	CAAC	780
DB	800	CAAGAA	CATT	CA	CGAC	ATCAT	CCCT	GC	ACAG	AGTGC	CCCG	AGCT	GGGG	GAGAT	GC	CAAC	859
QY	781	AAATA	AGAA	AGAT	CT	GAG	CTCT	CT	GGT	GGC	CTT	TTGG	CTT	AAAT	CT	GC	840
DB	860	AAATA	AGAA	AGAT	CT	GAG	CTCT	CT	GGT	GGC	CTT	TTGG	CTT	AAAT	CT	GC	919
QY	841	CAGT	CGCC	CT	ATT	TGT	CA	ACCT	CTT	TTGTT	TTCC	CGG	ACCT	TTGGT	GGC	AGT	900
DB	920	CAGT	CGCC	CT	ATT	TGT	CA	ACCT	CTT	TTGTT	TTCC	CGG	ACCT	TTGGT	GGC	AGT	979
QY	901	AGAGG	CAGT	GGCG	ATTT	TGA	CAG	CCAC	AT	ACCT	GTGG	TGC	AT	GC	CA	CT	960
DB	980	AGAGG	CAGT	GGCG	ATTT	TGA	CAG	CCAC	AT	ACCT	GTGG	TGC	AT	GC	CA	CT	1039
QY	961	GACG	GAAT	CCGT	CTGT	GTAT	CT	CGT	CTTT	TCG	AC	AGAT	TA	ACCC	CAG	CA	1020
DB	1040	GACG	GAAT	CCGT	CTGT	GTAT	CT	CGT	CTTT	TCG	AC	AGAT	TA	ACCC	CAG	CA	1099
QY	1021	GAGC	CAG	AGCA	AC	CAG	CT	AC	GGC	AGCC	CAC	AT	CA	GAG	AGT	TC	1080
DB	1100	GAGC	CAG	AGCA	AC	CAG	CT	AC	GGC	AGCC	CAC	AT	CA	GAG	AGT	TC	1159
QY	1081	TCT	CT	CACT	CA	CGT	CT	GT	TTT	TCG	T	GA	TGTT	TG	CAC	AC	1140
DB	1160	TC	CT	CACT	CA	CGT	CT	GT	TTT	TCG	T	GA	TGTT	TG	CAC	AC	1219
QY	1141	GAT	AGA	CAT	CAT	CGA	GT	G	ACT	TTG	CTT	G	AGAA	CT	CT	GT	1200
DB	1220	GAT	AGA	CAT	CAT	CGA	GT	G	ACT	TTG	CTT	G	AGAA	CT	CT	GT	1279
QY	1201	CTT	CT	CTCT	CTT	CC	AGT	TC	CA	GT	TC	AGT	GGG	CG	CA	TC	1260
DB	1280	CTT	CT	CTCT	CTT	CC	AGT	TC	CA	GT	TC	AGT	GGG	CG	CA	TC	1339
QY	1261	ACT	GA	AGAA	AA	ACC	TTG	TC	CTT	GGC	CC	AG	CT	CT	GT	CT	1320
DB	1340	ACT	GA	AGAA	AA	ACC	TTG	TC	CTT	GGC	CC	AG	CT	CT	GT	CT	1399
QY	1321	CAG	CT	CTGT	CT	CT	AC	CT	AC	CT	GGG	GT	GC	AC	GGT	CG	1380
DB	1400	CAG	CT	CTGT	CT	CT	AC	CT	AC	CT	GGG	GT	GC	AC	GGT	CG	1459
QY	1381	CCT	GG	CGG	CT	TTG	G	AG	AA	TCC	AC	AT	GGT	CG	CC	AT	1440
DB	1460	CCT	GG	CGG	CT	TTG	G	AG	AA	TCC	AC	AT	GGT	CG	CC	AT	1519
QY	1441	GA	AC	CAG	AA	AG	AG	AT	G	AGT	AGT	AGT	AGT	AGT	AGT	AGT	1500
DB	1520	GA	AC	CAG	AA	AG	AG	AT	G	AGT	AGT	AGT	AGT	AGT	AGT	AGT	1579
QY	1501	AG	CAT	GT	CT	CC	GA	AG	AGT	GC	AC	AG	AT	GC	GT	GA	1560
DB	1580	AG	CAT	GT	CT	CC	GA	AG	AGT	GC	AC	AG	AT	GC	GT	GA	1639
QY	1561	AGG	C	CGG	AG	CG	CAT	GGG	C	ACT	GC	AG	GA	CGT	CAGT	CAG	1620
DB	1640	AGG	C	CGG	AG	CG	CAT	GGG	C	ACT	GC	AG	GA	CGT	CAGT	CAG	1699
QY	1621	CT	CT	CCCT	CT	CC	CA	T	CGT	AT	TTT	TTT	TTT	TTT	TTT	TTT	1680
DB	1700	CT	CT	CCCT	CT	CC	CA	T	CGT	AT	TTT	TTT	TTT	TTT	TTT	TTT	1759
QY	1681	GG	CC	TT	GA	TTT	AA	AG	TTT	TC	GT	CT	AA	T	CT	AG	1740

1880	Db		CATACCCCTGCTCAGAAAAACCCAAAGACACAGCTGCTCACGGTTGACGTTGTGTCC	1939
1861	Qy		TCTCCCTCGGACAAATCTCTCTTGGAAACAAAGGACTGCAGCTGTGCCATCGCGCTCG	1920
1940	Db		TCTCCCTCGGACAAATCTCTCTTGGAAACAAAGGACTGCAGCTGTGCCATCGCGCTCG	1999
1921	Qy		GTACCCCTGACAGCAGGCGACAGACTCTCTGTGCCCTTCAATCGCTCTTAAGAATCAA	1980
2000	Db		GTACCCCTGACAGCAGGCGACAGACTCTCTGTGCCCTTCAATCGCTCTTAAGAATCAA	2059
1981	Qy		CAGTTTAAACTCGGCTTCTCTTGTGTTGCTTCCCACTCACATGGCCGTACAAAGAGATG	2040
2060	Db		CAGTTTAAACTCGGCTTCTCTTGTGTTGCTTCCCACTCACATGGCCGTACAAAGAGATG	2119
2041	Qy		GAGCCCGGTGGCTCTTAAATTTCCCTTGTGCCAGGTTGCGAAACCATCTACTCCAC	2100
2120	Db		GAGCCCGGTGGCTCTTAAATTTCCCTTGTGCCAGGTTGCGAAACCATCTACTCCAC	2179
2101	Qy		ACATGACGAGGCGGTGCACGCTCAGCCCGGAGTCCCGTTTCACTGTAGAGAACGGA	2160
2180	Db		ACATGACGAGGCGGTGCACGCTCAGCCCGGAGTCCCGTTTCACTGTAGAGAACGGA	2239
2161	Qy		GACCTGTGACACAGCAGGCTGACAGATGCACAGAAATCTCCGTAGAAAGTTTGGTTG	2220
2240	Db		GACCTGTGACACAGCAGGCTGACAGATGCACAGAAATCTCCGTAGAAAGTTTGGTTG	2299
2221	Qy		AAATGCCCGGGGCGACCAACTGACATGGTTGAATGATAGCAATTTCACTCGGTTCTC	2280
2300	Db		AAATGCCCGGGGCGACCAACTGACATGGTTGAATGATAGCAATTTCACTCGGTTCTC	2359
2281	Qy		CTAGATCTGACGAAGCTGTCAAGTTCTCACCCCCACCGTGTATATACATGAGCTTAACTTTT	2340
2360	Db		CTAGATCTGACGAAGCTGTCAAGTTCTCACCCCCACCGTGTATATACATGAGCTTAACTTTT	2419
2341	Qy		TTAAATTTGTACAAAGCGCATCTCCAGATTCAGACCCCTGCCGCTGACTTTTCTGAA	2400
2420	Db		TTAAATTTGTACAAAGCGCATCTCCAGATTCAGACCCCTGCCGCTGACTTTTCTGAA	2479
2401	Qy		GGCTTGCTTTTCCCTCGCCTTTCTCGAAGTGCATTTAGACGAGTCACATGGAGCATCC	2460
2480	Db		GGCTTGCTTTTCCCTCGCCTTTCTCGAAGTGCATTTAGACGAGTCACATGGAGCATCC	2539
2461	Qy		TAACTTTGCAATTTAGTTTTTACGTGAACGTGAAGCTTTAAGTCTCATCAGCATTTCTAA	2520
2540	Db		TAACTTTGCAATTTAGTTTTTACGTGAACGTGAAGCTTTAAGTCTCATCAGCATTTCTAA	2599
2521	Qy		TGCCAGGTGTGCTAGGCTAACTTTTGAAGTAGATATATTACCTGGTTCTGCTACTCTTA	2580
2600	Db		TGCCAGGTGTGCTAGGCTAACTTTTGAAGTAGATATATTACCTGGTTCTGCTACTCTTA	2659
2581	Qy		GTCAATACTCTGCGGTACAGTAATGTACAGAAATGTACTCCCTCCCAACACCAT	2640
2660	Db		GTCAATACTCTGCGGTACAGTAATGTACAGAAATGTACTCCCTCCCAACACCAT	2719
2641	Qy		ACGATAAGCAAGACATTTTATACGATACCGAGTCACATATGTGGTCTCCCTCGAATA	2700
2720	Db		ACGATAAGCAAGACATTTTATACGATACCGAGTCACATATGTGGTCTCCCTCGAATA	2779
2701	Qy		ACGCATTCGAAATCCATCGAGTGTATATTTTCTAAGTTTTTGGAAAGCAGGTTTTTT	2760
2780	Db		ACGCATTCGAAATCCATCGAGTGTATATTTTCTAAGTTTTTGGAAAGCAGGTTTTTT	2839
2761	Qy		CCTTTAAAAAATTTATAGACAGCGTTCACATAATGTATTTAGTCAGAAATTCCTAGATGA	2820
2840	Db		CCTTTAAAAAATTTATAGACAGCGTTCACATAATGTATTTAGTCAGAAATTCCTAGATGA	2899
2821	Qy		AAGAACTTAAACAAAAAATTTTAAAGATATAAATATATGCTGTATATGTTATGTAAT	2880
2900	Db		AAGAACTTAAACAAAAAATTTTAAAGATATAAATATATGCTGTATATGTTATGTAAT	2959
2881	Qy		TTATTTAGGCTATAATACATTTTCCCTATTTTTCGATTTTCAATTAATAAATGCTCTTAATACA	2940

FEATURES